

Alexander Refsum Jensenius. Using motiongrams in the study of musical gestures (pp 499-502)

Published in:

Proceedings of the 2006 International Computer Music Conference, November 6 - 11, 2006, New Orleans, USA (ICMC2006). ISBN 0-9713192-4-3, 9780971319240



Using Motiongrams in the Study of Musical Gestures

Alexander Refsum Jensenius

Musical Gestures Group, Department of Musicology, University of Oslo
Input Devices and Music Interaction Lab, Schulich School of Music, McGill University
a.r.jensenius@imv.uio.no

Abstract

Navigating hours of video material is often time-consuming, and traditional keyframe displays are not particularly useful when studying single-shot studio recordings of music-related movement. This paper presents the idea of motiongrams and how we use such displays in our studies of dancers' free movements to music.

1 Introduction

In our current research on music-related gestures (i.e. physical movement) and music, such as mimicry of sound-producing gestures (Godøy, Haga, and Jensenius 2006) and free movement to music (Casciato, Jensenius, and Wanderley 2005), we have found the need for tools to visualize movement-related information from video material.

Developing video summaries and systems for navigating large video databases have become increasingly more important as collections of digital video grow (Shipman, Girgensohn, and Wilcox 2005). However, most systems we have come across focus on detecting and displaying new camera shots and content of scenes (e.g. people, animals and places) rather than movement-related information. The former is useful when working with video material with many different shots and scenes, but our studio recordings have only one shot, no camera movement and only one subject. Thus a traditional timeline display of video frames is not so useful for our needs (see Figure 1 made with Metadata Hootenanny¹), since most of the frames may look similar unless we happen to sample a salient posture by chance.

Trying to improve such timeline displays, we have experimented with making keyframe displays based on detecting salient postures (Figure 2) (Jensenius, Godøy, and Wanderley 2005). The stored frames in such a non-linear display reveal more interesting movement qualities than a purely time-sampled display, but it is more difficult to get a sense of temporal development when frames are stored at an uneven rate.

Data	TL 0	TL 1	TL 2	TL 3	TL 4	TL 5	TL 6	TL 7	TL 8	TL 9
1 00:00:00.00										
501 00:00:20.00										
1001 00:00:40.00										
1501 00:01:00.00										
2001 00:01:20.00										
2501 00:01:40.00										
3001 00:02:00.00										
3501 00:02:20.00										
4001 00:02:40.00										
4501 00:03:00.00										

Figure 1: A traditional timeline display is not very useful when studying free dance movement to music.

Another problem with both the timeline and posture-based displays is that they do not reveal much of the actual *motion* in the sequence.

In the audio domain we are used to working with spectrograms, which give an idea about changes in spectral content and loudness, and how the music evolves over time. We have been interested in creating something similar for video material, so that we can more easily visualize motion over time.

This paper presents the idea of creating *motiongrams* from video material, and how such displays can be used when studying musical gestures.

2 Motiongrams

A popular way of visualizing motion is by calculating the difference between consecutive frames in a video stream. Such *motion images* (Figure 3) show only the pixels in the im-

¹ Available from <http://www.applesolutions.com/bantha/MHguide.html>



Figure 2: Extracting salient frames from a dance sequence based on changes in quantity of motion.

age that change between frames, and is thus a good indication of a person's movement. We particularly find *motion history images* to be useful, made with feedback of previous frames, since they create motion traces that may resemble our short term memory. However, such "trails" cannot be too long, typically in the range between 2 to 10 seconds dependent on the level of motion, otherwise the image would only look blurred. The challenge is thus to find a way of representing motion in longer sequences.



Figure 3: Different motion images (left to right): raw frame-difference, with smoothing and threshold functions, with edge detection, and with a feedback function creating a motion history image.

One approach is to calculate *quantity of motion* by summing up the active pixels in a motion image and plotting the value over time. Such graphs give some idea about overall motion qualities, but remove all spatial information about where in the image (hence in the body) the motion is taking place. We have therefore been interested in trying to create a display that would show the quantity of motion over time, while at the same time preserving some of the spatial information.

Our current approach is based on collapsing a matrix of size $M \times N$ into $1 \times N$ and $M \times 1$ matrices, by averaging over the columns and rows respectively. Plotting these 1 pixel wide stripes against time results in what we call *motiongrams*. Figure 4 shows vertical and horizontal motiongrams of a dance sequence. Notice how visible the movement of the dancer's hands and head are, and how it is possible to follow the trajectories they make over time.

Usually we prefer to work with grayscale motion images since colours often seem to be more of a distraction than help. In the case of our dance recordings, however, we have found it very useful to use colours since the dancers used differently

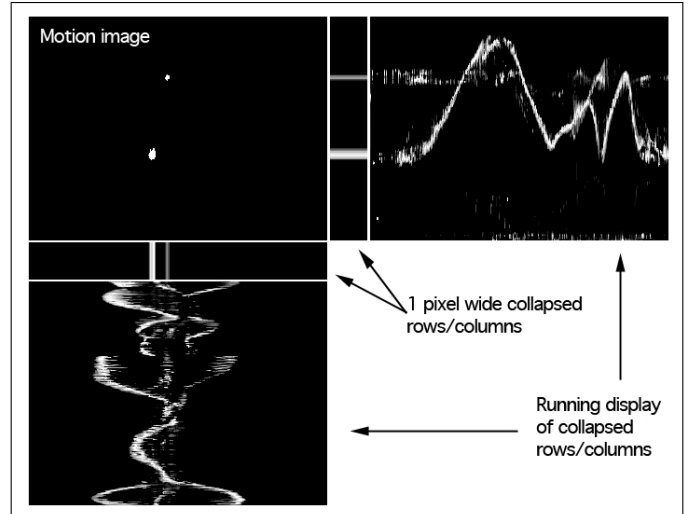


Figure 4: Motion image, collapsed motion image, and running collapsed motion images (motiongrams) of a 10 second long dance sequence.

coloured gloves and socks (originally meant for doing colour blob tracking). This makes it possible to visually separate the different "streams" in the motiongrams by following the colours of the dancer's gloves, feet and head (Figure 6).

The motiongrams have been created using Max/MSP/Jitter and modules from the Musical Gestures Toolbox². Figure 5 shows an overview of the process which starts by cleaning up the original video by automatically cropping the video based on maximum contraction, and increasing the brightness, contrast and saturation to get a clear image with strong colours and contrast. Next we find the colour motion image by subtracting consecutive frames, and doing some noise reduction to clean it up (with blur and threshold functions). Finally matrix reduction is done by calculating the mean of each row of the image³ and adding these frame slices together to the motiongram image.

Figure 6 shows a motiongram of a 5 minute sequence of dance movements to music, where the dancer was asked

²Beta version available from <http://musicalgestures.uio.no>

³Using `xray.jit.mean` by Wesley Smith, available from <http://www.mat.ucsb.edu/~whsmith/xray.html>

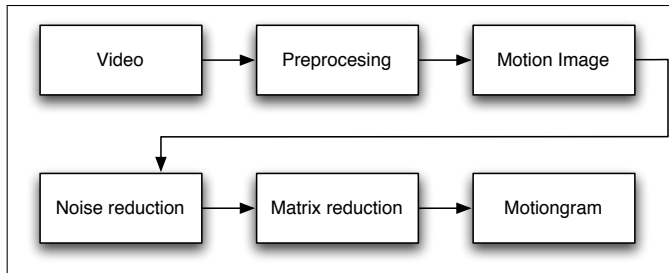


Figure 5: Overview of process.

to move freely to five different musical examples each repeated three times. Although quite compressed, the motiongram clearly reveals the structure of the sequence and the levels of movement. It is possible to see that the dancer was mainly moving her arms up and down in the beginning of the sequence, and it is also easy to spot when she was almost standing still between the music excerpts. Having a display like this makes it possible to quickly navigate in the material, and start video playback from various positions in the file by pointing at the corresponding location in the motiongram.

In addition to using the motiongrams for navigational purposes, we also find that they are useful for comparative studies. Figure 7 shows a 40 second extract of free dance movements by three different dancers, the same musical stimuli repeated three times. The display makes it possible to compare the performances, and study differences and similarities in hand and head movement.

3 Discussion

This paper has presented how we create *motiongrams* from videos by collapsing video frames into 1 pixel wide matrices which are plotted against time. The resulting images display the level and location of motion in the video, and makes it easy to follow trajectories over time.

One of the big challenges in our study of relationships between gestures and sounds, is to develop methods and tools for navigating and visualizing audio and video content (and even sensor information) in a way that opens for comparison. We find motiongrams useful when it comes to quickly getting an overview of large amounts of video material, in the same way as we are used to working with spectrograms in the audio domain. They are, indeed, crude reductions of the original video material, but they also manage to grasp some movement qualities that are otherwise difficult to represent.

The method is purely based on a transformation of motion images, with no analysis taking place (e.g. colour-tracking or gesture recognition). This makes the approach stable, and works well with all sorts of video material where the camera

is standing still. The method is even quite useful for getting an overview of the structure and content in movies and music videos, although rapid changes in shots, camera movements and zooming are obviously much more prominent in such motiongrams than in the ones based on a fixed camera recording.

Future work includes:

- Creating three dimensional motiongrams to account for both horizontal and vertical motion.
- Creating multidimensional motiongrams from multiple camera recordings.
- Improving processing speed when creating motiongrams in Max/MSP, and implement the algorithm in Matlab.
- Creating combined displays for video, audio and sensor data.

Acknowledgments

Thanks to Rolf Inge Godøy and Marcelo M. Wanderley for valuable feedback and support. This research is funded by the Norwegian Research Council.

References

- Casciato, C., A. R. Jensenius, and M. M. Wanderley (2005). Studying free dance movement to music. In *Proceedings of ESCOM 2005 Performance Matters! Conference, Porto, Portugal*.
- Godøy, R. I., E. Haga, and A. R. Jensenius (2006). Playing ‘air instruments’: Mimicry of sound-producing gestures by novices and experts. In S. Gibet, N. Courty, and J.-F. Kamp (Eds.), *Gesture in Human-Computer Interaction and Simulation: 6th International Gesture Workshop, GW 2005, Berder Island, France, May 18-20, 2005, Revised Selected Papers*, Volume 3881/2006, pp. 256–267. Springer-Verlag GmbH.
- Jensenius, A. R., R. I. Godøy, and M. M. Wanderley (2005). Developing tools for studying musical gestures within the Max/MSP/Jitter environment. In *Proceedings of the International Music Computer Conference, Barcelona, 4-10 September, 2005*, pp. 282–285. Barcelona, Spain: International Computer Music Association.
- Shipman, F., A. Girgensohn, and L. Wilcox (2005). Hyper-video expression: Experiences with hyper-hitchcock. In *Sixteenth ACM Conference on Hypertext and Hypermedia, September 6, 2005*.

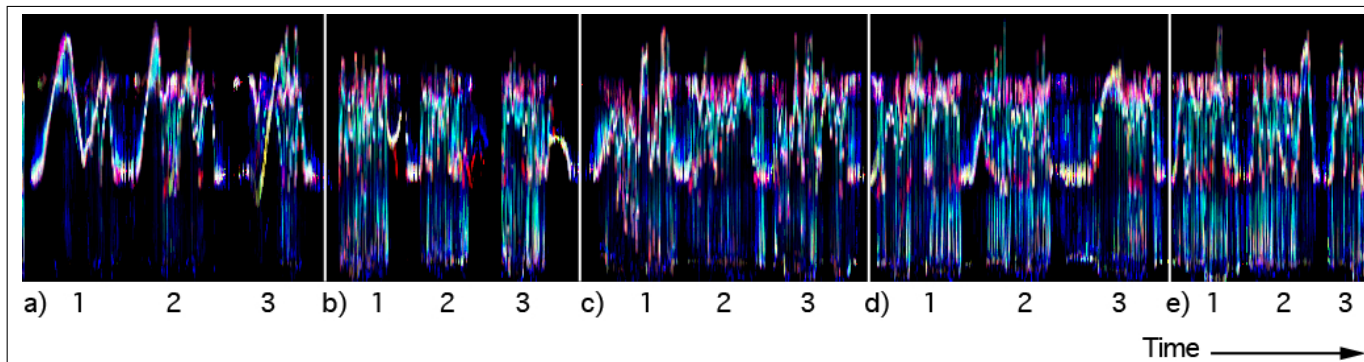


Figure 6: Motiongram of a 5 minute video of free dance movements to music. The dancer moved to 5 different musical excerpts (a-e) and each excerpt was repeated three times (1-3). Although quite rough, it is easy to see differences in quantity of motion and similarities in upward/downward patterns between the sequences.

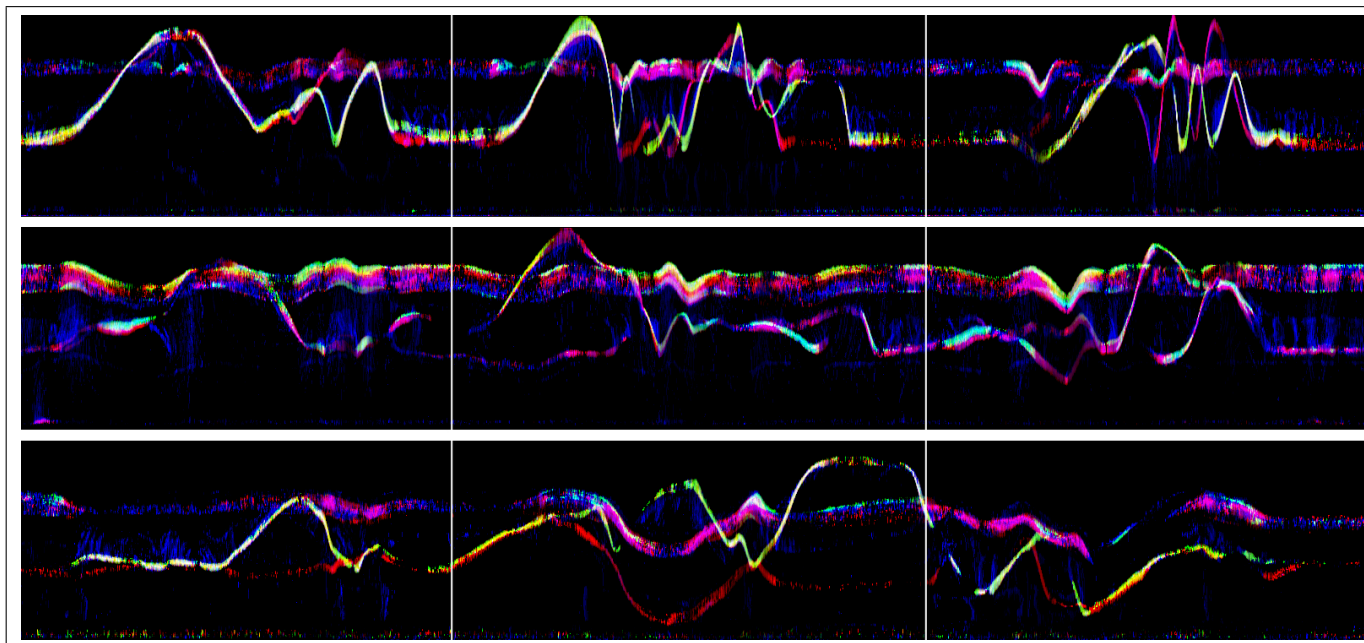


Figure 7: Motiongrams of three dancers doing free movements to the same excerpts of music repeated three times (total time displayed is approximately 40 seconds). It is easy to follow the two hands (yellow and red) and head (pink due to saturation), as well as the body (appears as blue stripes due to the blue background).